Computer Vision Project 1 - MyAutoPano

Kaushik Kavuri Subrahmanya
Robotics Engineering
Worcester Polytechnic Institute
Worcester, Massachusetts 01609
Email: ksubrahmanya@wpi.edu
Using 2 Late Days

Butchi Adari Venkatesh
Robotics Engineering
Worcester Polytechnic Institute
Worcester, Massachusetts 01609
Email: badari@wpi.edu
Using 2 Late Days

***Abstract*—In this project, we aim to recreate the Structure from Motion (SfM) procedure where we reconstruct a 3D scene and simultaneously obtain the camera poses of a monocular camera w.r.t. the given scene. In SfM, we create the entire rigid structure from a set of images with different viewpoints, or equivalently, a camera in motion. The output of the steps are shown and discussed below.**

I. Dataset for Classical SfM

We use 5 images of Unity Hall at WPI taken using a Samsung S22 Ultra's primary camera at f/1.8 aperture, ISO 50, and 1/500 sec shutter speed. The camera is calibrated, and the images are distortion-corrected and resized to 800*600px. SIFT keypoints and descriptors are provided and keypoint matching between each image and its successive images is also provided.

Fig. 1: Feature Descriptors of the sample image

A. Estimating Fundamental Matrix

The Fundamental Matrix is a matrix that describes the relationship between corresponding points in two images of a scene taken from different viewpoints. The matrix satisfies the equation $x'^T Fx = 0$. The F matrix is a 3x3 matrix that is obtained by solving the homogeneous linear system with 9 unknowns:

$$[x_i' y_i' 1] \begin{bmatrix} f_{11} & f_{21} & f_{31} \\ f_{12} & f_{22} & f_{32} \\ f_{13} & f_{23} & f_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0$$

We need 8 points to solve this system of equations. Due to noise in the correspondences, the estimated F matrix can be of rank 3. Since the rank needs to be 2, the issue is corrected by setting the last(smallest) singular value of the estimated F matrix to zero.

B. Match Outlier Rejection via RANSAC

As the point correspondences of SIFT could contain noise, it may have several outliers. We use RANSAC algorithm to remove these outliers to obtain a better estimate of the F matrix. RANSAC results in that F matrix with the highest number of inliers.

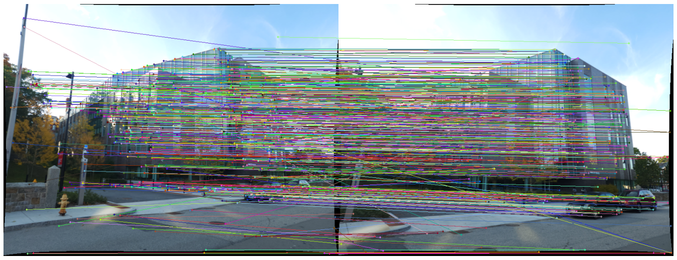

Fig. 2: Feature Matching Before RANSAC

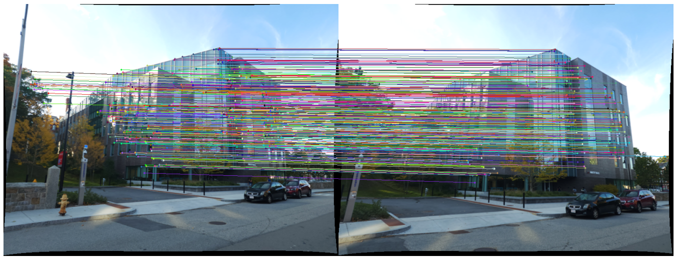

Fig. 3: Feature Matching After RANSAC

C. Estimate Essential Matrix from Fundamental Matrix

Since the F matrix was calculated using epipolar constrains, the relative camera poses between the two images can also be calculated using the F matrix. Relative camera poses can be computed using the Essential Matrix, E which is 3×3 matrix, that satisfies the equation $E = K^T FK$, where K is the camera calibration/intrinsic matrix.

Similar to F matrix, the noise in K matrix might result in the singular values of E not being $(1, 1, 0)$. This can be corrected by svd of E and reconstructing it as $E = U \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T$

D. Estimate Camera Pose from Essential Matrix

The pose of a camera is the rotation (Roll, Pitch, Yaw) and translation (X, Y, Z) of the camera with respect to the world. We can obtain the four camera pose configurations

$(C_1, R_1), (C_2, R_2), (C_3, R_3)$ and (C_4, R_4) using E where C is the camera center and R is the rotation matrix.

For $E = UDV^T$ and

$$W = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

the 4 configurations are written by:

$$C_1 = U(:,3)$$
$$R_1 = UWV^T$$

$$C_2 = -U(:,3)$$
$$R_2 = UWV^T$$

$$C_3 = U(:,3)$$
$$R_3 = UWV^TV^T$$

$$C_4 = -U(:,3)$$
$$R_4 = UWV^TV^T$$

E. Triangulation Check for Cheirality Condition

The Cheirality condition states that the reconstructed points must be in front of the cameras. To check for this condition, we triangulate the 3D points of two camera poses using linear least squares to check the sign of the depth Z in the camera coordinate system with respect to camera center. A 3D point X is in front of the camera iff:

$$\mathbf{r}_3(\mathbf{X} - \mathbf{C}) > 0$$

The best camera configuration, (C, R, X) is the one that produces the maximum number of points satisfying the cheirality condition. Using te 4 camera pose and the linearly triangulated points from above, we can disambiguate the camera pose.

1) Non-Linear Triangulation: Given the camera poses and the linearly triangulated points, the locations of the 3D points that minimizes the reprojection error can be further refined. The minimization formula is of the form:
$\min_x \sum_{j=1,2} \Big((u^j - P_1^j T\tilde{X}/P_3^{jT}X)^2 + (v^j - P_2^{jT}\tilde{X}/P_3^{jT}X)^2\Big)$

F. Perspective-n-Points, Linear Camera Pose Estimation

We have 3D points in the world, their 2D projections in the image and the intrinsic parameter K. Now, the 6 DOF camera pose can be estimated using linear least squares. Given 2D-3D correspondences, $X \leftrightarrow x$, and the intrinsic parameter K, we estimate the camera pose using linear least squares.

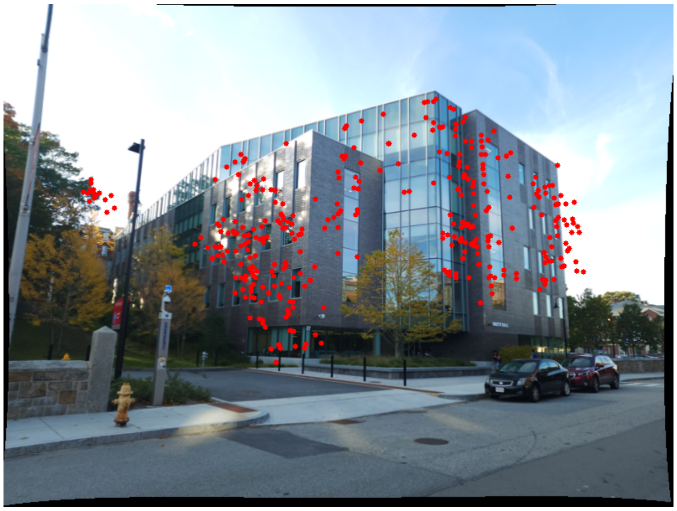

Fig. 4: Reprojected Points after Linear Triangulation - 1

Fig. 5: Reprojected Points after Linear Triangulation - 2

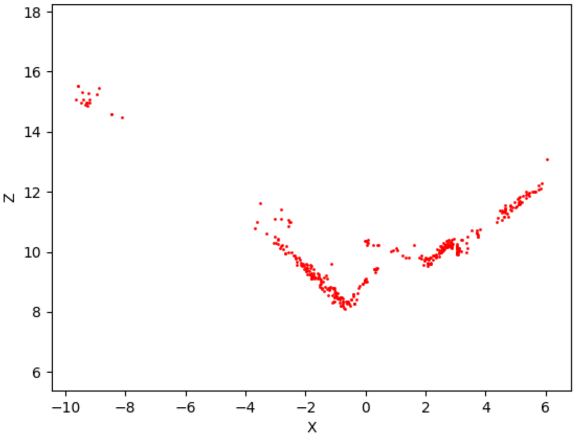

Fig. 6: Points Plot of Non Linear

G. PnP RANSAC

The PnP method above could have errors as as there could be outliers in the given set of point correspondences. To overcome this, we use RANSAC to help eliminate outliers.

H. Nonlinear PnP

We can refine the camera pose that minimizes reprojection error. In this optimization step, we convert the rotation matrix into quarternion form as it is a better choice to enforce orthogonality of the rotation matrix.

I. Bundle Adjustment

Using our initialized camera poses and 3D points, we refine them further by minimizing reprojection error using bundle adjustment. Bundle adjustment refines camera poses and 3D points simultaneously by minimizing the reprojection error over $C^I_{i_i=1}$, $q^I_{i_i=1}$, and $X^I_{i_i=1}$.

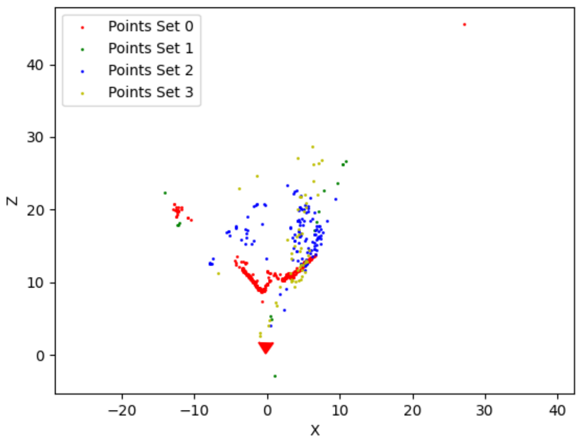

Fig. 7: Plot Points with Camera Poses

Method	Img2	Img3	Img4	Img5
Linear Triangulation	4.01	20.49	53.3	89.88
Non Linear Triangulation	3.87	20.7	52.01	56.02
Linear PnP (RANSAC)	-	69.28	90.8	137.3
Non Linear PnP	-	84.5	134.2	135.3

TABLE I: Re-projection error

II. Conclusion

We therefore implemented the components of Structure from Motion following the classical approach to this problem

Acknowledgment

The authors would like to thank the professor for the instructions and knowledge imparted via the course.